Forum Analisis Statistik
Juni 2024, 4 (1): 31-38

Clustering of Regencies in Kalimantan Barat Based On Community Welfare Indicators

Wahyudio Shaney Fikri Hariadi[1*], Shantika Martha[2], Desa Ayu Natalia[3]

[1,2]Universitas Tanjungpura, Jl. Prof. Dr. H. Hadari Nawawi, Pontianak;
[3]BPS Provinsi Kalimantan Barat, Jl. Sutan Syahrir, Pontianak;
*Penulis Korespondensi. e-mail: h1091211030@student.untan.ac.id

ABSTRACT

Welfare is the capacity to meet all life's necessities to live decently, healthily, and productively. The objective of this study is to classify districts/cities in Kalimantan Barat Province according to indicators of community welfare. The data utilized include the percentage of impoverished individuals, gross regional domestic product, average years of schooling, expected years of schooling, and life expectancy. The procedure in this study commenced with a descriptive analysis of each variable, followed by the formation of an Euclidean distance matrix. After that, clustering with the centroid linkage method, in this case using the value of the silhouette coefficient to determine the optimal number of clusters formed. The clusters' outcomes are visible in the dendrogram. Cluster traits are identified based on the mean value of each variable within each cluster. The findings of the analysis indicate that Kota Pontianak enjoys high welfare, while the other thirteen districts are classified as having low welfare levels.

Keywords: cluster, centroid linkage, euclidean, welfare, Kalimantan Barat

INTRODUCTION

Sustainable Development Goals (SDGs) is a program that aims to promote better economic development based on the foundations of people, planet, prosperity, peace, and partnership (Wahyuningsih, 2017). Welfare is a measure of a society that has been in a state of prosperity (Mulia & Saputra, 2020). A person is more prosperous if they possess a higher level of education because they are already established in terms of work and income (Salsabila et al., 2022). Welfare is the capacity to meet all life's necessities to live decently, healthily, and productively (Ilmiah et al., 2022). Indonesia is still trying to improve the welfare of its people (Shrestha & Coxhead, 2018). Kalimantan Barat is one of the provinces in Indonesia with a population of 5.623 million with a male population of 2.887 million and a female population of 2.736 million and with a productive age population aged 15-64 years in Kalimantan Barat of 2.922 million, this can be an asset for Kalimantan Barat Province to be able to improve the welfare of its people. Where the productive age population is the economic support of a region (BPS, 2023).

Various ways are planned and carried out by the region to increase growth for the region itself, with the hope that it will have a positive impact on the welfare of its people (Sihite, 2022). The indicators of community welfare cover various aspects, including poverty, education, income, employment, living standards and consumption patterns, as well as aspects of housing and the environment (Nabilah et al., 2024). Gross regional domestic product is used to see the level of welfare in terms of the economy as a whole (Ilmiah et al., 2022). Meanwhile, unemployment and poverty are barriers to enhancing community welfare. According to BPS, the percentage of poor people in Kalimantan Barat ranks 23rd among the 34 provinces (BPS, 2024). Life Expectancy (AHH) is a tool used to assess government

performance to improve the welfare of the community from the aspect of health (Valiant Kevin et al., 2022). AHH is used to determine the factors that influence the improvement of people's welfare in the region (Herawaty Bangun, 2019). Therefore, it is essential to categorize the welfare characteristics of Kalimantan Barat Province to determine whether Kalimantan Barat is in a group or cluster with high, low, or moderate community welfare.

This research uses data from the Central Bureau of Statistics (BPS). The data used are indicators of community welfare from each district / city in Kalimantan Barat in 2023, specifically the percentage of poor people (PPM), gross regional domestic product (GRDP), average years of schooling (RLS), expected years of schooling (HLS), and life expectancy (AHH). These five indicators have a direct relationship with community welfare.

METHODOLOGY

Cluster Analysis

Cluster analysis is a multivariate analysis technique that groups one object with another. Objects that are in one cluster have the same relative proximity to other objects (Pasha Pratama et al., 2023). Cluster analysis has two methods, namely hierarchical and non-hierarchical. Cluster methods that are organized and stratified based on the similarity of properties between objects are called hierarchical method. Some methods in calculating the closest distance in the hierarchical method are single linkage, complete linkage, average linkage, ward, and centroid linkage (Silvi,2018). The non-hierarchical method begins with determining the number of clusters to minimize the objective function or optimal criteria. Next, one cluster that has the closest distance is merged with another cluster (Widyawati et al., 2020).

Data Standardization

Data standardization is done using the following z-score formula:

$$Z_i = \frac{x_i - \bar{x}}{s}, s \neq 0 \quad (1)$$

Where: Z_i represented variable standardization, x_i represented data to i, $\bar{x}$ represented overall average of data for each variable, and s represented standard deviation.

Euclidean Distance Calculation

The distance measure between objects that is often used to calculate dissimilarity is the euclidean distance. The euclidean distance formula is (Purwanty et al., 2024):

$$d_{ik} = \sqrt{\sum_{j=1}^{p} (X_{ij} - X_{kj})^2} \quad (2)$$

Where: d_{ik} represented distance from i-th object to k-th object, X_{ij} represented the i-th object on the j-th variable, X_{kj} represented the k-th object in the j-th variable, j represented 1,2,…….,p and p represented number of variables.

Centroid Linkage

Centroid linkage is a clustering phase based on the distance between centroids. This method is effective for reducing variance within clusters (Pratama et al., 2023). The advantage of using this method is that the outliers do not significantly affect the cluster analysis tested (Silvi, 2018). The following centroid formula is used:

$$\overline{X} = \frac{N_1\overline{x}_1 + N_2\overline{x}_2}{N_1 + N_2} \quad (3)$$

Where: $\overline{X}$ represented centroid value, $\overline{x}_i$ represented distance from object value, N_i represented number of objects, and i represented 1,2.

Silhouette Coefficient

This technique is a clustering evaluation method that integrates both cohesion and separation approaches. To deliver insights into the quality of clustering outcomes during the clustering process, the silhouette of each cluster and even the whole cluster of the performance of a clustering algorithm can be calculated. Here is the silhouette coefficient formula:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i)} \quad (4)$$

Where: $s(i)$ represented silhouette coefficient value, $a(i)$ represented average distance between i and all other points in the same cluster as i, $b(i)$ represented average distance between i and all points in the nearest cluster that does not contain i.

RESULT AND DISCUSSION

The secondary data used in this study are indicators of the welfare status of Kalimantan Barat Province in 2023. These include five variables: the Percentage of Poor Population (PPM), Gross Regional Domestic Product (GRDP), Average Years of Schooling (RLS), Expected Years of Schooling (HLS), and Life Expectancy (AHH).

From the data on community welfare indicators, a hierarchical analysis was conducted using the centroid linkage method. The following will be shown the data description of each variables:

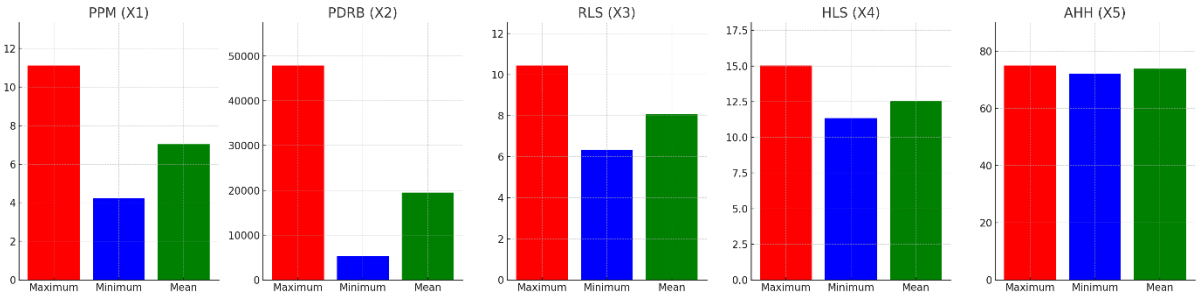

Figure 1. Data Description

Source: Central Bureau of Statistics (BPS) of Kalimantan Barat, the data is being processed.

Based on Figure 1, it can be seen that the amount of data for all variables is 14 districts/cities in Kalimantan Barat. With the largest average value on the gross regional domestic product variable (X_2) and the smallest average value on the percentage of poor people variable (X_1).

Data Standardization

Because the units of each variable are different, we must standardize the data using z-score before testing. There are 14 districts/city objects used, namely Sambas (A), Bengkayang (B), Landak (C), Mempawah (D), Sanggau (E), Ketapang (F), Sintang (G), Kapuas Hulu (H), Sekadau (I), Melawi (J), Kayong Utara (K), Kubu Raya (L), Kota Pontianak (M), and Kota Singkawang (N).

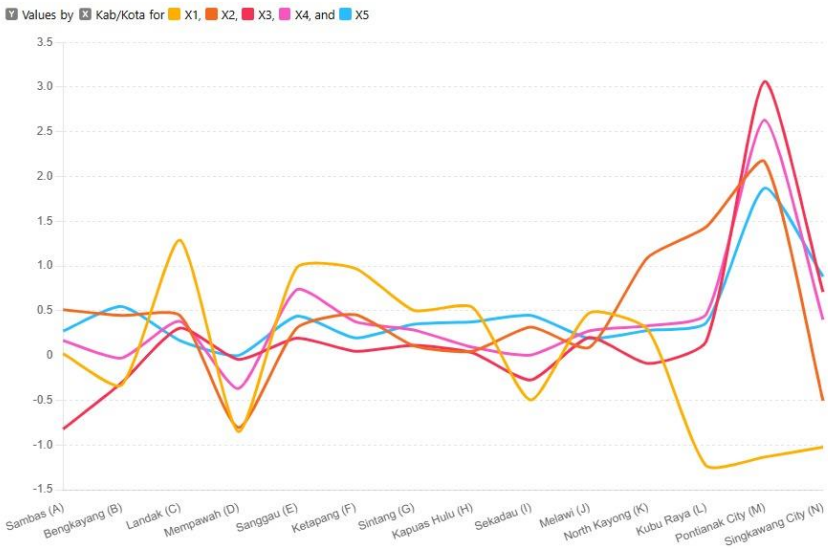

Figure 2. Data Standardization

Source: Central Bureau of Statistics (BPS) of Kalimantan Barat, the data is being processed

Distance Matrix Formation

The first step in forming a distance matrix is to calculate the distance between A and B using the data in Figure 2 as follows:

$$d_{A,B} = \sqrt{\left(0.02102 - (-0.33031)\right)^2 + \left(0.51243 - (-0.63727)\right)^2 + \cdots + (0.2739 - 0.5478)^2} = 1.456 \quad (5)$$

After that, repeat the calculation in the same way until you get a distance matrix of size 14×14

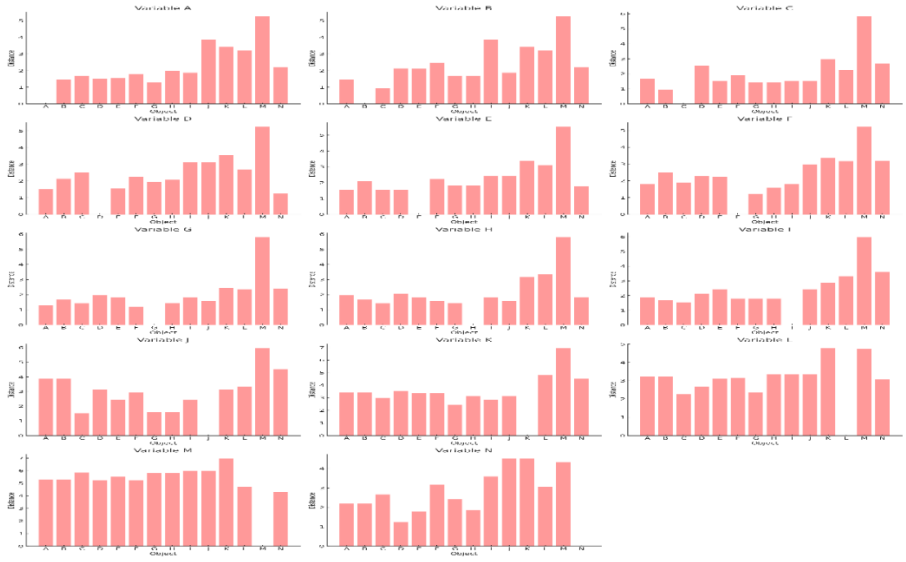

Figure 3. Distance Matrix Formation

Central Bureau of Statistics (BPS) of Kalimantan Barat, the data is being processed

After the distance matrix is formed, then the next step is to determine the smallest / closest distance, namely at a distance of I to B with a value of 0.558 (Figure 3).

Distance Value with Centroid Linkage Method

At this stage, the centroid value is calculated based on the minimum value in the distance matrix (I,B) using the formula:

$$\bar{X}_{I,B} = \frac{5.9 \times 1 + 6.28 \times 1}{1 + 1} = \frac{12.18}{2} = 6.09 \quad (6)$$

Repeat the calculation of the centroid value of each object until the results are obtained in Figure 4 below:

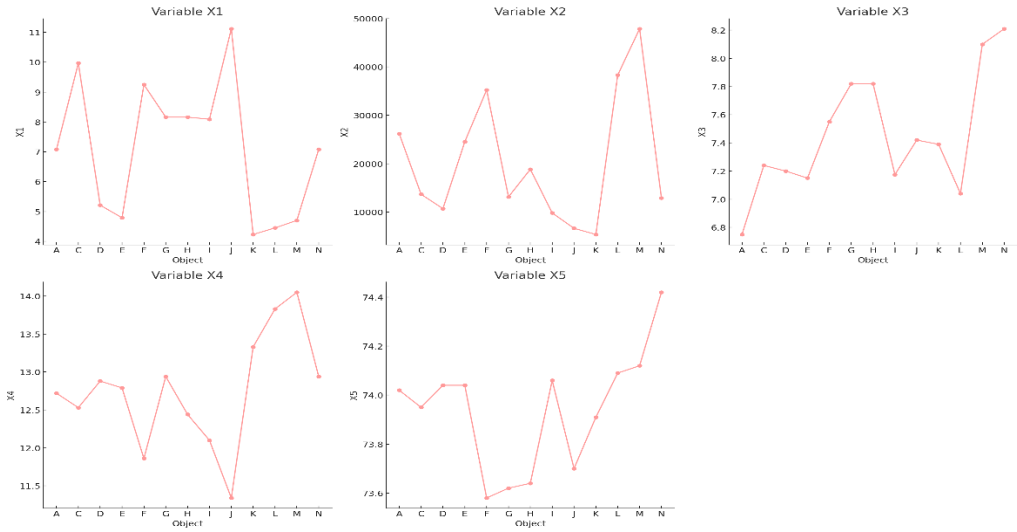

Figure 4. Distance Value with Centroid Linkage

Source: Central Bureau of Statistics (BPS) of Kalimantan Barat, the data is being processed

After calculating the centroid value, then calculate the euclidean distance using the data in Figure 4. The euclidean distance matrix using the centroid linkage method can be seen in Table 2.

Silhouette Coefficient Value

For this step, the silhouette coefficient value is calculated to determine the optimal cluster created.

Table 1. Silhouette Coefficient Value

No	Cluster	Silhouette Value
1	K2	0.27038
2	K3	0.22574
3	K4	0.23277
4	K5	0.23536
5	K6	0.23926
6	K7	0.19801
7	K8	0.15409
8	K9	0.14009

For the largest silhouette coefficient value, there is K2 with a silhouette coefficient value of 0.27038, it can be said that the optimal cluster to be formed is 2 clusters.

Table 2. Distance Matrix

	M	(N,I,B,D,E,F,A) KE (H,G,C,J,L,M)
M	0	
(N,I,B,D,E,F,A) KE (H,G,C,J,L,M)	5.47248	0

Dendrogram Formation

Based on Figure 5, there are 14 districts/cities used in the formation. Two clusters were formed using the centroid linkage method based on indicators of community welfare in Kalimantan Barat Province in 2023. Cluster one consists of Kota Pontianak, cluster two is Kayong Utara, Kubu Raya, Melawi, Ketapang, Kota Singkawang, Sanggau, Sambas, Mempawah, Bengkayang, Sekadau, Kapuas Hulu, Landak, and Sintang.

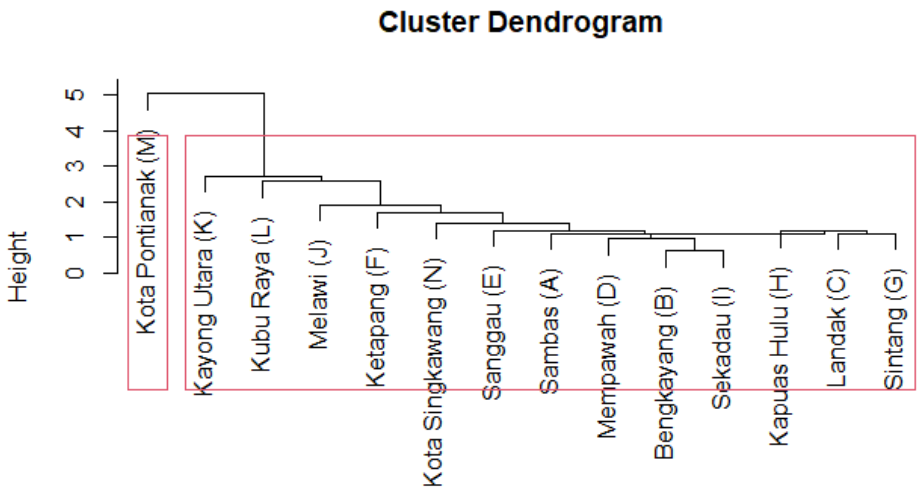

Figure 5. Cluster Dendrogram of Community Welfare by District/City in Kalimantan Barat 2023

Source: Central Bureau of Statistics (BPS) of Kalimantan Barat, the data is being processed

Cluster Characteristics

Each cluster member obtained is analyzed to determine the characteristics of each variable in each cluster.

Table 3. Characteristics of Cluster 1

Member	Variable				
	X_1	X_2	X_3	X_4	X_5
Kota Pontianak	4.45	47,885.24	10.45	15.04	75.07
Average Cluster 1	**4.45**	**47,885.24**	**10.45**	**15.04**	**75.07**

Table 4. Characteristics of Cluster 2

Member	Variable				
	X_1	X_2	X_3	X_4	X_5
Kayong Utara	9.13	5,375.79	6.35	12.12	72.17
Kubu Raya	4.23	38,323.80	7.04	13.89	73.27
Melawi	11.12	6,644.40	7.42	11.34	73.70
Ketapang	9.25	35,249.16	7.55	11.96	73.58

Member	Variable				
	X_1	X_2	X_3	X_4	X_5
Kota Singkawang	4.70	12,899.30	8.21	12.94	74.42
Sanggau	4.79	24,529.32	7.44	11.87	74.01
Sambas	7.08	26,202.78	6.75	12.72	74.02
Mempawah	5.21	10,714.01	7.20	12.88	74.04
Bengkayang	6.28	11,146.26	7.22	12.16	74.20
Sekadau	5.90	8,506.34	7.13	11.92	73.93
Kapuas Hulu	8.16	13,140.86	7.82	12.20	73.33
Landak	9.97	13,684.30	7.24	12.53	73.95
Sintang	8.18	254,613.30	7.64	12.30	74.07
Average Cluster 2	**7.2115**	**19,999.20**	**7.48**	**12.5331**	**73.7954**

Based on Table 3, and Table 4 with the centroid linkage method, it is known that there are differences between the two clusters. Cluster 1 is distinguished by having a lower average percentage of impoverished individuals compared to cluster 2, yet it exhibits higher average values for gross regional domestic product, years of schooling (both average and expected), and life expectancy.

CONCLUSION AND RECOMMENDATION

Based on the findings from the analysis conducted on clustering districts/cities in Kalimantan Barat Province using the centroid linkage method, it can be concluded that among the 14 districts/cities, two clusters were formed based on indicators of community welfare in 2023. Kota Pontianak (cluster 1) is an area characterized by a high level of welfare, whereas Kayong Utara, Kubu Raya, Melawi, Ketapang, Kota Singkawang, Sanggau, Sambas, Mempawah, Bengkayang, Sekadau, Kapuas Hulu, Landak, and Sintang (cluster 2) have a low level of welfare.

For this research suggestion, local governments can look at the cluster results and the variables/indicators used as a reference to see the development of local communities, especially in areas in cluster 2 with the highest percentage of impoverished individuals, whereas the gross regional domestic product, average years of schooling, expected years of schooling, and life expectancy are at their lowest. Perhaps the local government can make cluster 2 a priority to improve community welfare and make cluster one a benchmark for community welfare.

REFERENCES

Central Bureau of Statistics (BPS) of Kalimantan Barat. (2023). *Indikator Kesejahteraan Rakyat Provinsi Kalimantan Barat 2023 (Vol. 15)*. Pontianak: Badan Pusat Statistik Provinsi Kalimantan Barat

Central Bureau of Statistics (BPS) of Kalimantan Barat. (2024). *Provinsi Kalimantan Barat Dalam Angka 2024 (Vol. 41)*. Pontianak: Badan Pusat Statistik Provinsi Kalimantan Barat

Herawaty Bangun, R. (2019). Analisis Determinan Angka Harapan Hidup Kabupaten Mandailing Natal (Life Expectations Determinants Analysis In Mandailing Natal Regency). 4(3). https://ojs.unpkediri.ac.id/index.php/akuntansi/article/view/13257

Ilmiah, J., Erudisi, E. &, Afri Mulia, R., & Putri, R. P. (2022). JIEE: Pengaruh Tingkat Kemiskinan dan Produk Domestik Regional Bruto Terhadap Kesejahteraan Masyarakat. 2, 2022–2044. https://journal.unespadang.ac.id/JIEE/article/view/93

Mulia, R. A., & Saputra, N. (2020). Analisis Faktor-Faktor Yang Mempengaruhi Kesejahteraan Masyarakat Kota Padang. https://doi.org/http://dx.doi.org/10.24014/jel.v11i1.10069

Nabilah, N. A., Perdana, H., & Sulistianingsih, E. (2024). *Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Kesejahteraan Masyarakat dengan Algoritma K-Means++*. In Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster) (Vol. 13, Issue 3).

Pasha Pratama, P., Debataraja, N. N., & Aprizkiyandari, S. (2023). *Pengelompokkan Kabupaten di Provinsi Kalimantan Timur Berdasarkan Tingkat Kemiskinan dengan Metode Centroid Linkage*. In Buletin Ilmiah Math. Stat. dan Terapannya (BIMASTER) (Vol. 12, Issue 4).

Purwanty Purwanty, Shantika Martha, & Siti Aprizkiyandari. (2024). Pengelompokan Kabupaten/Kota di Kalimantan Barat Berdasarkan Faktor-Faktor Kemiskinan Menggunakan Metode Dbscan. 13, No 1. https://doi.org/https://dx.doi.org/10.26418/bbimst.v13i1.74047

Rosaria Sihite. (2022). Analisis Pengaruh Pendapatan Perkapita, Jumlah Konsumsi dan Pertumbuhan Ekonomi di Kabupaten/Kota Provinsi Kalimantan Tengah. https://doi.org/https://doi.org/10.52300/jepp.v2i1.4433

Salsabila, Q., Lukluli, A., & Cahyono, H. (2022). Kondisi Sosial Ekonomi Terhadap Kesejahteraan Keluarga di Jawa Timur. https://doi.org/https://doi.org/10.26740/independent.v2n3.p169-177

Shrestha, R., & Coxhead, I. (2018). Can Indonesia Secure a Development Dividend from Its Resource Export Boom? *Bulletin of Indonesian Economic Studies*, *54*(1), 1–24. https://doi.org/10.1080/00074918.2018.1458702

Silvi, R. (2018). Analisis Cluster dengan Data Outlier Menggunakan Centroid Linkage dan K-Means Clustering untuk Pengelompokkan Indikator HIV/AIDS di Indonesia. *Jurnal Matematika "MANTIK," 4*(1), 22–31. https://doi.org/10.15642/mantik.2018.4.1.22-31

Valiant Kevin, A., Bhinadi, A., & Syari'udin, A. (2022). Pengaruh PDRB, Angka Harapan Hidup, dan Rata Rata Lama Sekolah Terhadap Kemiskinan di Kabupaten/Kota Provinsi Jawa Tengah Tahun 2013-2021. *SIBATIK JOURNAL: Jurnal Ilmiah Bidang Sosial, Ekonomi, Budaya, Teknologi, Dan Pendidikan*, *1*(12), 2959–2968. https://doi.org/10.54443/sibatik.v1i12.482

Wahyuningsih. (2017). Millenium Develompent Goals (MDGs) dan Sustainable Development Goals (SDGs) Dalam Kesejahteraan Sosial. In *Bisma Jurnal Bisnis dan Manajemen* (Vol. 11, Issue 3). https://doi.org/https://doi.org/10.19184/bisma.v11i3.6479

Widyawati, W., Saptomo, W. L. Y., & Utami, Y. R. W. (2020). Penerapan Agglomerative Hierarchical Clustering Untuk Segmentasi Pelanggan. *Jurnal Ilmiah SINUS*, *18*(1), 75. https://doi.org/10.30646/sinus.v18i1.448